

**Трофимов С.П., Довбуш П.П.**

**МНОГОФУНКЦИОНАЛЬНАЯ ONLINE-СИСТЕМА ХРАНЕНИЯ ГЕНЕАЛОГИЧЕСКОЙ ИНФОРМАЦИИ И ЕЁ ИСПОЛЬЗОВАНИЕ В МУЛЬТИМЕДИЙНЫХ ОБРАЗОВАТЕЛЬНЫХ ПРОЕКТАХ**

*tsp61@mail.ru*

*ГОУ ВПО УГТУ-УПИ*

*г. Екатеринбург*

*Разработана система ввода, хранения и представления генеалогической информации через Интернет. Реализовано экспортирование и импортирование базы данных системы из внутреннего формата в международный стандарт GEDCOM. Предложены способы статистической обработки данных, позволяющие экстраполировать генеалогическое дерево на будущие поколения.*

*The system of input, storage and representation of the genealogic information through the Internet is developed. Export and importation between internal format and international standard GEDCOM is realized. Ways of the statistical data processing for extrapolating a genealogic tree on the future generations are offered.*

До начала компьютерной эры вся информация создавалась, хранилась и передавалась на бумажных носителях. В наши дни они дублируются мультимедийными электронными носителями. Вместе с тем, некоторые виды информации невозможно обрабатывать вне коммуникационной среды, стандартом которой сейчас стала сеть Интернет. Примером такой области является генеалогия. Образование населения в данной сфере является важной государственной задачей. Одному человеку невозможно собрать воедино отдельные веточки и листья генеалогического дерева (ГД). Это под силу грамотному обществу, владеющему современными технологиями и желающему совместными усилиями построить структуру генеалогической ткани, которая сплачивает это общество.

Если судить по публикациям в Интернете и по проходящим в разных городах России генеалогическим выставкам, за последние два-три года количество людей в России, интересующихся генеалогией, возросло в десятки раз.

Родословные данные являются в разных странах важной частью национальных программ по народосбережению. Например, в Исландии подобные данные хранятся с 10-го века. При этом фиксируются данные о наследственных и других заболеваниях, профессиональная деятельность, имущественное состояние.

Генеалогическая информация может быть использована также и в мультимедийных образовательных проектах. Генеалогические деревья великих династий могут быть использованы на уроках истории или литературы для достижения различных дидактических целей:

- дать учащимся наглядное представление о правящих династиях в рамках заданного исторического периода;
- познакомить учащихся с предками или потомками исторических личностей или выдающихся деятелей культуры;

- обучение поисково-исследовательской деятельности (найти и выстроить недостающие звенья).

Построение генеалогического дерева своей семьи может быть предложено учащимся в курсе обществознания или граждановедения при изучении темы «Семья». Использование мультимедийных технологий делает эту работу для учащихся более интересной и продуктивной.

При исследовании существующих средств для работы с ГД было найдено и проанализировано более 40 сайтов, в основном англоязычных. Большая часть из них содержит статьи о ручном (не компьютерном) способе поиска информации и составлении генеалогических деревьев, а также статьи, посвященные использованию различных программ, позволяющих работать с ГД и экспортировать результаты работы в GEDCOM – стандартизированный формат хранения ГД.

Было найдено несколько полноценных online-систем, позволяющих искать информацию (в основном по [www.ancestry.com](http://www.ancestry.com)). Однако в большинстве случаев добавление новой информации в этих системах осуществляется администраторами сайтов, а не самими пользователями. Результаты поиска генеалогической информации по этим системам в основном представляются пользователю в виде текстовом виде без отображения всех связей или в виде одноуровневого дерева.

Из отечественных сайтов можно отметить [www.bird.dn.ru](http://www.bird.dn.ru), который позволяет вводить и отображать генеалогическую информацию. Но организация ввода представляется не очень удобной.

Большая часть найденных web-страниц представляют собой персональные странички пользователей. Генеалогические деревья на таких страницах чаще всего представлены в текстовом виде, иногда выполнены вручную в системе Microsoft WORD или с использованием программ для работы с ГД. Не было найдено ни одной реализации, предоставляющей возможность просматривать результаты поиска или добавлять информацию в естественном для пользователя виде, то есть в виде дерева.

Некоторые сайты предлагают пользователю вводить информацию в очень специфичном, отличном от GEDCOM, формате, описанием которого располагают только программисты-разработчики.

В результате анализа существующих решений была отмечена важная черта успешного проекта в данной области – поддержка импорта/экспорта данных в стандартный формат GEDCOM и организация обмена генеалогическими данными. Условие совместимости со стандартами стало, по существу, основной идеологией дальнейшей работы.

Задача создания полноценной online-системы хранения генеалогических деревьев и работы с ними достаточно сложна. Такая система должна содержать:

- базу данных,
- аналитическую и статистическую обработку содержимого баз данных,
- простой и сложный поиск по базам данных;

- а также допускать:
- возможность добавления информации администратором,
- возможность просмотра имеющейся информации,
- возможность просмотра и добавления информации клиентом,
- обмен информацией между клиентами,
- конвертирование содержимого баз данных в стандартизированные форматы обмена генеалогическими данными.

При разработке системы основное внимание было уделено удобству пользовательского интерфейса, а также возможности импорта генеалогических данных по стандарту GEDCOM XML.

При создании системы было решено ориентироваться на бесплатные технологии, что существенно снизит затраты на разработку, развертывание и поддержку системы. Были использованы многие современные компьютерные технологии: реляционные базы данных и хранимые процедуры – MySQL версии 5.1, новейший подход к построению интерактивных пользовательских интерфейсов веб-приложений – Ajax, экономный протокол общения между клиентом и сервером – JSON. С целью снижения требований к серверу и каналу передачи данных основная нагрузка по представлению информации ложится на браузер клиента («толстый» клиент).

Для облегчения процесса разработки и отладки было решено разделить систему на следующие, практически независимые части (см. Рис.):

- Сервер: Обеспечение хранения данных и доступа к ним.
- Сервер (взаимодействие с клиентом): Обработка запросов клиентов и предоставление им информации.
- Клиент (взаимодействие с сервером): Обеспечение связи с сервером и предоставление данных в понятном серверу формате; обработка ответов сервера и ошибок.
- Клиент: Представление пользователю данных в естественном виде.

Подобная модульность и иерархичность позволяет облегчить разработку системы и выявление неполадок. При этом облегчается и развитие проекта – модуль с неудовлетворяющей функциональностью может быть заменен независимо от всей системы (что, например, было сделано при смене формата данных для обмена между сервером и клиентом – переходе от XML к JSON).

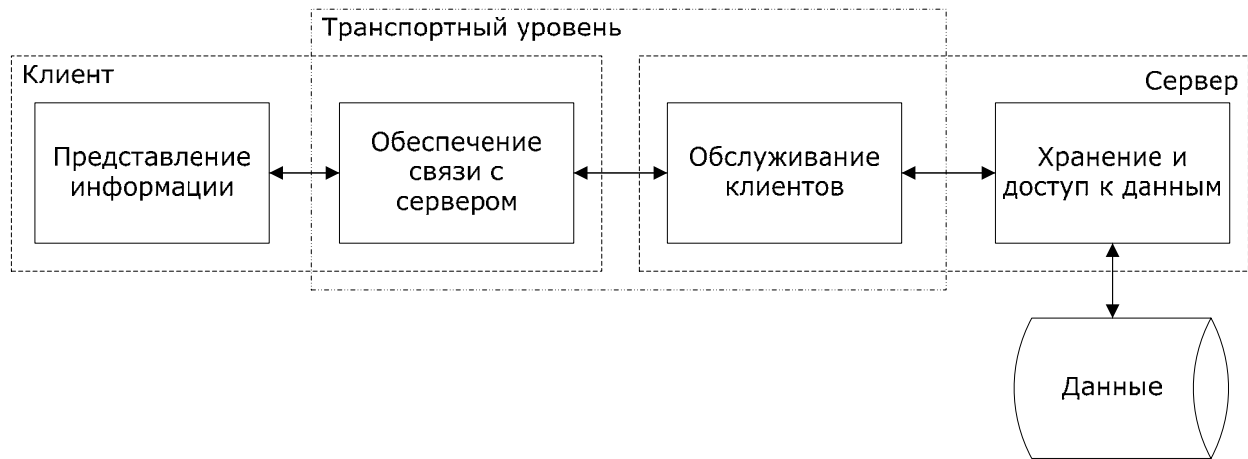


Рис. Схема взаимодействия компонентов системы

Использование новейшей методологии Ajax позволило нам резко снизить нагрузку на сервер, повысить скорость реакции системы на действие пользователя и удобство работы.

**Ajax** (от англ. *Asynchronous JavaScript and XML* – «асинхронный JavaScript и XML») – это подход к построению интерактивных пользовательских интерфейсов веб-приложений. При использовании Ajax веб-страница не перезагружается полностью в ответ на каждое действие пользователя. Вместо этого с веб-сервера догружаются только нужные пользователю данные. Сервер передаёт не готовый HTML-код, а только данные; а HTML-элементы создаются исходя из этих данных, с использованием методов модификации DOM (DOM – объектная модель документа). При этом в качестве формата передачи данных обычно используются XML или JSON.

При построении клиент-серверного взаимодействия было опробовано несколько вариантов и выбран наиболее экономичный из них – JSON. Проведен анализ существующих форматов динамически генерируемой графики и выбран наиболее подходящий и перспективный из них.

После первоначальной загрузки страницы клиент получает от сервера только запрошенную информацию. В этом состоит коренное отличие от традиционного подхода, при котором вся страница загружается заново. Например, при запросе информации о человеке получаем только его ФИО и связанные с ним данные. Скрипты на JavaScript представляют полученные данные в запрошенном пользователем формате: текстовом, в виде редактируемой формы или в виде дерева. Также, для удобства пользователя, при помощи JavaScript, добавлены некоторые функции помогающие пользователю (например, сокрытие или отображение (не) нужной ему в данный момент части документа).

Для **хранения** древовидных структур (графов) придумано много способов. Но большинство из них рассчитаны на графы простой структуры, а листья генеалогического дерева содержат большое количество информации, не требуемой при выполнении операций над графом (например, поиск родственных связей). Поэтому в работе предлагается сделать двухуровневый способ хранения информации – все данные о человеке хранить на SQL сервере, а для поиска информации использовать индекс своего формата (GTI). От использования

только SQL-сервера было решено отказаться, так как обход дерева реализуется рекурсивным алгоритмом, в котором информацию о следующем шаге мы получаем на шаге текущем. В этом случае даже незначительное количество, например десять, последовательных SQL запросов требуют огромного количества времени.

**Отображение** генеалогической информации в виде дерева (наиболее важная и интересная составная часть создаваемой системы) является сложной теоретической задачей. При исследовании существующих алгоритмов автоматического отображения деревьев нами были отмечены работы John Q. Walker II, в которых описан алгоритм позиционирования узлов для отображения деревьев общего вида. Но фактически, генеалогическое дерево, с точки зрения теории графов, является не деревом, а сетью, так как у каждого узла-человека есть два родителя и, кроме того, существует возможность наличия нескольких супружеских связей. Это не позволяет нам использовать большинство существующих алгоритмов. На данный момент задача находится в процессе решения.

#### Описание стандарта GEDCOM

GEDCOM (от англ. Genealogical Data Communications) – формат обмена генеалогическими данными.

Файл GEDCOM – это открытый текст, содержащий информацию о людях и связях между ними. Структура файла GEDCOM утверждена мировым стандартом, последняя версия которого – 5.5 от января 1995 года.

Большинство современных генеалогических компьютерных программ поддерживает импорт/экспорт данных в формате GEDCOM.

В декабре 2002 года была выпущена бета-версия стандарта GEDCOM 6.0. В шестой версии в качестве формата хранения данных используется XML. По поддерживаемым возможностям новая версия стандарта не содержит практически ничего нового относительно версии 5.5. Это – просто другой формат хранения той же информации в файле. Кроме бета-версии стандарта GEDCOM 6.0 существует несколько альтернативных предложений по хранению генеалогических данных в XML, например GedML.

Разработчики стандарта GEDCOM 6.0 XML предлагают файл XSLT преобразования для преобразования GEDCOM файлов версии 5.5 в 6.0. Работать с XML несколько проще, однако в ходе исследования вопроса не было найдено ни одной программы, поддерживающей импорт/экспорт данных в формате XML.

#### Различия между традиционным GEDCOM и XML GEDCOM

##### Цели разработки стандарта XML GEDCOM:

- Облегчить обмен генеалогическими данными в выразительном структурированном формате.
- Позволить обрабатывать и отображать данные при помощи веб-средств.

Традиционный GEDCOM разработан 15 лет назад. Технологии, использование и осмысление генеалогических данных изменились с тех пор. Спецификация GEDCOM 6.0 основана на более современных подходах. Обратная со-

вместимость с предыдущими версиями стандарта не поддерживается. Формат содержит почти те же данные, что и традиционный GEDCOM, но его структура более понятная и менее двусмысленная. Некоторые несущественные и малоиспользуемые возможности традиционного GEDCOM были исключены.

Несмотря на то, что новый стандарт отличается как синтаксисом, так и логической структурой, он все равно считается эволюцией традиционного GEDCOM.

В традиционном GEDCOM связи между записями двунаправленные, что конечно, избыточно и излишне. В XML GEDCOM связи однонаправленные.

Главное достоинство XML – в его расширяемости. Формат GEDCOM XML – это словарь для генеалогических данных. Спецификой XML является возможность его расширения без потери поддержки обратной совместимости – программы, не понимающие новых тэгов, просто и безболезненно пропустят их.

В стандарте GEDCOM использовались различные кодировки текста (как известно, для русского языка этих кодировок четыре), что вносило некоторую путаницу. XML GEDCOM (да и XML вообще) использует универсальную кодировку – Unicode.

Несмотря на все преимущества нового стандарта, он еще не утвержден официально и, из-за отсутствия поддерживающего его программного обеспечения, было решено использовать GEDCOM версии 5.5.

Структура и пример GEDCOM файла

В табл. 1 приведены только основные тэги GEDCOM и их XML аналоги, формирующие основной каркас документа.

Таблица 1  
Основные тэги GEDCOM.

GEDCOM 5.5	GEDCOM 6.0 XML	Описание
HEAD TRLR	HeaderRec -	Заголовок. Содержит информацию о файле. Признак конца файла
FAM HUSB WIFE CHIL MARR	FamilyRec HusbFath WifeMoth Child BasedOn	Семья муж/отец жена/мать дети Свадьба (событие)
INDI NAME SEX EVENT  DATE  PLAC	IndividualRec IndivName Gender PersInfo  Date  Place	Человек Имя (в формате « Имя Отчество /Фамилия/ ») Пол Событие Дата Место

В табл. 2 приведен пример файла GEDCOM версии 5.5, содержащего минимально возможную информацию.

Таблица 2  
Пример файла GEDCOM

0 HEAD 1 SOUR DGEN 2 VERS 0.10 2 NAME DPP Genealogy 2 CORP Paul Dovbush 3 ADDR http://gen.krorm.ru/ 1 DEST ANY 1 FILE test.ged 1 GEDC 2 VERS 5.5 2 FORM Lineage-Linked 1 CHAR ANSEL 1 LANG Russian	Заголовок – информация о файле      Используемая версия GEDCOM  Кодировка Язык
0 @I1@ INDI 1 NAME Адам // 1 SEX M 1 FAMS @F1@	Человек №1 – Адам, пол мужской
0 @I2@ INDI 1 NAME Ева // 1 SEX F 1 FAMS @F1@	Человек №2 – Ева, пол женский
0 @I3@ INDI 1 NAME Каин // 1 SEX M 1 FAMC @F1@	Человек №3 – Каин, пол мужской
0 @F1@ FAM 1 HUSB @I1@ 1 WIFE @I2@ 1 MARR 1 CHIL @I3@	Семья №1 – муж – Адам, жена Ева, сын – Каин.
0 TRLR	Признак конца файла

Генеалогическое дерево может содержать значительный объем статистической информации. Мы планируем использовать эту информацию для экстраполяции ГД на будущие поколения и для восстановления наиболее вероятного развития жизни родоначальников этого ГД. Известный русский философ Н. Федоров в своей “философии общего дела” определил “воскресение отцов, как главную задачу христианской цивилизации”.

Статистические параметры ГД мы разделяем на две группы.

Первая группа используется для повышения наглядности генетической информации. К этому, например, можно отнести следующие возможности:

- история семейных фамилий на протяжении всего периода ГД;
- география перемещения семейств в исторической перспективе;
- распределение по месяцам в виде круговых диаграмм различных событий: даты рождения, смерти, свадьбы ;
- демографическая пирамида – показывает соотношение между полами в зависимости от времени.

Вторая группа параметров касается распределений и корреляционных свойств, присущих ГД. Из них мы выделяем:

- плотности распределения возраста мужчин и женщин в момент заключения брака;
- плотность распределения возраста родителей при появлении первого ребенка;
- распределение времени от первого ребенка до второго.

Сюда мы также относим плотности распределения двумерных случайных величин:

- возраст обоих родителей при заключении брака;
- возраст обоих родителей при появлении первого ребенка;
- возраст одного из родителей при появлении первого и второго ребенка;
- возраст одного из родителей при появлении первого и последующее количество детей;
- плотность временного периода между детьми.

Эту информацию мы используем для моделирования динамики ГД. Для моделирования одномерных и двумерных случайных величин с заданными плотностями мы используем метод обратных функций. Если для системы случайных величин известна ковариационная матрица, то для их генерации мы используем метод линейного преобразования.

Хорошо заполненное дерево позволяет получить выборочные оценки для плотностей и ковариаций. Это позволяет экстраполировать свойственную именно для данного дерева динамику развития.

Другой подход при моделировании состоит в том, что мы произвольно задаем теоретические числовые и функциональные характеристики абстрактного дерева и затем моделируем его развитие на отдаленное будущее. При этом «выросшее» дерево должно удовлетворять ряду необходимых условий, например не разрастаться и не вырождаться.

Демонстрационная версия системы временно доступна по адресу <http://gen.krorm.ru>